

Centro de Informatización Universitaria

Facultad 1

Universidad de las Ciencias Informáticas



Trabajo de Diploma para optar por el título de Ingeniero
en Ciencias Informáticas

Título: Propuesta de solución para la integración del sistema de noticias
de la Universidad de las Ciencias Informáticas a la Red Social
Universitaria

Autora: Beatriz E. Osoria Tamayo

Tutores: Ing. Miguel Jaeger Rodríguez Lazo

Ing. Nayla Socarras Monzón

Ciudad de la Habana, Junio 2012

“Año 54 de la Revolución”



"En Cuba nadie ha hecho tanto en tan poco tiempo"

Fidel Castro Ruz

Declaración de autoría

Declaración de autoría

Declaro que soy el único autor de este trabajo y autorizo al Centro de Informatización Universitaria de la Universidad de las Ciencias Informáticas, para que haga el uso que estime pertinente con este trabajo.

Para que así conste firmo la presente a los ____ días del mes de _____ del año ____2012_____.

Autora: Beatriz Emilia Osoria Tamayo

Tutor: Ing. Nayla Socarras Monzón

Tutor: Ing. Miguel Jaeger Rodriguez Lazo

Agradecimientos

Primeramente le agradezco a la Revolución por haberme dado la oportunidad de estudiar y llegar hasta el momento en el que me encuentro.

Gracias a la UCI (Universidad de las Ciencias Informáticas) por permitirme ser una más de sus estudiantes y por darme la oportunidad de formarme y graduarme de Ing. en Ciencias Informáticas.

Agradezco a mi compañero, amigo y novio Osmel por haberme comprendido en mis momentos de furia y estrés, por su paciencia y comprensión.

A mi nueva familia, por ser ese nuevo rincón de refugio cuando ya el estrés me estaba quitando las fuerzas.

A mis amigos, que sin ellos no hubiera podido lograrlo, Grethel, Chacón, Yisel, Yamí, Jessica, Arlenís, Gleidy, Zori.

Al grupo Infodanz, gracias por esas horas de relajación que siempre me dieron, los quiero mucho.

A todas aquellas personas que formaron parte de esta maravillosa etapa de mi vida que ahora culmina, que a lo largo de 5 años me han hecho disfrutar de su amistad, a los que se fueron y los que aún quedan, muchas gracias, nunca voy a olvidarlos.

A mis tutores por hacer posible la realización de esta tesis, por guiarme y ayudarme en esta etapa crucial de mi carrera, en especial a Miguel Jaeger Rodríguez Lazo.

A mi tía Consuelo, por ser esa persona que siempre me ha dado un consejo oportuno cuando menos lo quiero, pero cuando más lo necesito. Te quiero.

A lo más grande y hermoso que tengo en este mundo: mi madre, por ser esa persona que nunca se dio por vencida conmigo, por ser mi motor impulsor, por sacar fuerzas de donde no hay para seguir alentándome, por poner toda su vida y más en función de mi felicidad, por ser mi ejemplo de superación en la vida profesional y en esa larga fiesta de experiencias que es la vida, en fin por ser mi orgullo. Te agradezco tu dedicación, tus horas de desvelo, tu sabiduría y por qué no tus regaños, pues son los que me han llevado a ser la persona que soy hoy.

Por último pero no menos importante a mi voz de conciencia: mi abuela, que aún cuando físicamente no esté, la siento muy cerca de mí, dándome todo su apoyo incondicional y cuidándome desde arriba.

*A mi madre,
por todo el trabajo y sacrificio
y por ser mi motor impulsor.*

Resumen

El desarrollo de la presente investigación se centra en la propuesta de un mecanismo de integración basado en las tecnologías de la Web Semántica que permita la publicación de noticias en la red de forma estandarizada. La investigación aborda un análisis del estado del arte de los mecanismos de integración utilizados por algunas de los principales medios de comunicación con otras aplicaciones, incluidas las redes sociales. De la misma forma se realiza un análisis de los mecanismos de la Web Semántica empleados para la vinculación de datos y componentes utilizados para ese fin. Al finalizar la investigación se escoge el mecanismo más idóneo para resolver el problema científico del trabajo.

Palabras clave: integración, medio de comunicación, red social, Web Semántica.

Abstract

The development of this research focuses on the proposal for a mechanism of integration based on Semantic Web technologies to enable the publication of news on the network in a standardized way. The research addresses an analysis of the state of the art integration mechanisms used by some of the main means of communication with other applications, including social networks. Similarly, will be analyzing the Semantic Web mechanisms used for data binding and components used for this purpose. When the research is done we will provide the most appropriate mechanism to solve the scientific problem of labor.

Keywords: integration, media, social networking, Web Semantics.

Índice

Introducción	1
Capítulo 1: Fundamentación teórica	6
1.1. Introducción	6
1.2. Redes sociales	6
1.3. Sistema de Gestión de contenido (CMS)	8
1.3.1. Wordpress	8
1.3.2. Tecnologías de desarrollo.....	9
1.3.3. Herramientas de desarrollo.....	12
1.3.4. Sistemas Gestores de Base de Datos	14
1.4. Web Semántica	15
1.4.1. Mapa conceptual de la Web Semántica	18
1.5. Ontologías	19
1.5.1. Lenguajes de ontologías	20
1.5.2. Estándares de metadatos disponibles (Web Semántica).....	22
1.5.3. Estándares XML	22
1.6. Mecanismos de integración entre aplicaciones	24
1.6.1. RSS.....	24
1.6.2. API (<i>Application Programming Interface</i>).....	25
1.6.3. Linked Data (Datos vinculados).....	26
1.7. Diferencias entre las Web API y Linked data	27
Capítulo 2: Propuesta de solución	28
2.1. Introducción	28
2.2. Diagnóstico del campo de acción	28
2.3. Descripción de la propuesta	29
2.3.1. Linked data.....	29
2.3.2. Propuesta para exportar en formato NewsML.....	33
Capítulo 3: Validación de la propuesta	38
3.1. Introducción	38
3.2. Validación de la propuesta	38
3.2.1. Primer caso: BBC	38
3.2.2. Segundo caso: The New York Times.....	40
3.2.3. Tercer caso: DBpedia Mobile.....	40
3.2.4. Cuarto caso: Estadísticas INE	41

3.2.5. Quinto caso: GeoLinked Data.....	42
Conclusiones.....	44
Recomendaciones	45
Referencias bibliográficas.....	46
Bibliografía consultada.....	49

INTRODUCCIÓN

Desde el comienzo de la humanidad el hombre siempre tuvo la necesidad de comunicarse e interactuar con sus semejantes, compartir ideas, discutir opiniones; siempre ha tenido la tendencia de reunirse con personas que tengan sus mismos intereses y que a pesar de discrepar de su concepto e ideas, le aporten valores, conocimientos, criterios; creando entre ellos un lazo, un sistema de relaciones, una Red Social.

Con el decursar del tiempo, el desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC), el auge alcanzado por Internet y la creciente necesidad de comunicarse con muchas más personas y exponer sus puntos de vista, se da a conocer en 1995, como parte del avance científico de la época, la primera Red Social en línea.

El progreso de las redes sociales en Internet fue lento pero, la evolución tecnológica ha facilitado nuevas herramientas que permiten mejorar la comunicación virtual entre las personas; comenzando a desarrollarse cada vez más el mundo de las redes sociales en línea, y alcanzando reconocimiento con la aparición de algunas conocidas como Facebook, MySpace, Tribe, Friendster, entre otras.

Las redes sociales en línea en el ámbito mundial son una de las herramientas de comunicación más utilizadas de la llamada Web Semántica¹ o web 3.0 permitiendo que exista la interacción social y el intercambio digital entre personas, grupos o empresas que se identifican de alguna manera con los mismos intereses, necesidades o problemas.

La masificación y el auge que han adquirido las redes sociales en el mundo desarrollado han inducido a los medios de comunicación a centrar su atención en ellas, por supuesto, evaluando las posibles plataformas de difusión más aceptadas por el público. Aún cuando las redes sociales online no son medios de comunicación, sirven de canales que permiten la distribución de contenido de carácter noticioso.

Es tan fuerte la inserción de los medios de comunicación en las redes sociales online, que algunos de los más importantes a nivel mundial como el Washington Post, The New York Times, El Mundo.es, han logrado una fuerte integración con las más conocidas en estos momentos (Facebook, Twitter),

¹ Web semántica: Es un conjunto de actividades desarrolladas en el contexto de World Wide Web Consortium (W3C) tendente a la creación de tecnologías para publicar datos legibles por aplicaciones informáticas

incluyendo entre sus facilidades, que sus usuarios pueden compartir sus contenidos a través de sus páginas personales.

Cuba no ha tardado en insertarse en este nuevo mundo y ya cuenta con varios sistemas de gestión de noticias como Cubadebate, el periódico Granma digital, Juventud Rebelde entre otros, que al igual que los grandes a nivel mundial, se han integrado de manera exitosa con las redes sociales online más conocidas a nivel internacional.

Además de contar con estos medios de comunicación, también existen tres redes sociales internas en el país, la primera cuyo nombre es Neko, se encuentra en la Universidad de Villa Clara, RedSocial que está administrada por el Instituto Metalúrgico de Moa y la tercera tiene por nombre TSociales, red social de los trabajadores del país; aunque solo se puede acceder a ellas a través de la red nacional, es decir, vía Intranet, el hecho de que estén en funcionamiento significa un paso de avance.

Estas redes sociales todavía no están integradas a ninguno de los sistemas de gestión de noticias existentes en el país porque todavía no gozan de gran difusión, conocimiento por parte de la población, ni mecanismos que lo posibiliten.

Al igual que en el resto del país, en la Universidad de las Ciencias Informáticas (UCI), se busca un avance tecnológico tangible para migrar hacia la Web Semántica (web 3.0), utilizando las Nuevas Tecnologías de la Información y las Comunicaciones (NTIC), en la misma se creó un Centro de Informatización Universitaria (CENIA), dentro de los objetivos de este centro figura buscar las mejores y más apropiadas técnicas y mecanismos para lograr dicha migración exitosamente. Dentro de este centro existe una línea que se especializa en redes sociales, siendo su primera tarea realizar la Red Social Universitaria (RSU) de la UCI, un servicio dirigido a comunidades universitarias en el que se le permite a los usuarios intercambiar fotos, archivos, aplicaciones, mensajes de texto y otro tipo de contenidos online, pero la misma se quedó limitada en algunos aspectos.

En estos momentos el interés de la línea de trabajo es la realización de varios componentes que contribuyan a enriquecer la Red Social Universitaria implementada para la comunidad de la UCI. Dentro de los componentes desarrollados se encuentran una plataforma de creación de blogs personales y temáticos, un sistema encargado de consumir RSS² y un sistema de gestión de noticias. Actualmente los servicios que brinda este sistema actúan de forma independiente en el sitio, sin lograr

² RSS: son las siglas de Really Simple Syndication, un formato XML para syndicar o compartir contenido en la web. Se utiliza para difundir información actualizada frecuentemente a usuarios que se han suscrito a la fuente de contenidos.

que otros sistemas puedan interactuar con él, a pesar de constituir uno de los componentes de la todavía en construcción RSU. Constituye esta la limitación que presenta dicha aplicación; al no existir un mecanismo capaz de integrarla con la Red Social Universitaria, no provee ningún método para que otras aplicaciones sean capaces de recuperar la información expuesta; por lo que la información se encuentra duplicada en la red.

Debido al planteamiento anterior se expresa como **problema a resolver**: ¿Cómo lograr la integración del sistema de noticias Mella de la Universidad de las Ciencias Informática con la Red Social Universitaria, de manera que posibilite la exposición de los contenidos de forma estandarizada? Como consecuencia el **objeto de estudio** de la presente investigación es la Web Semántica, y el **campo de acción** en el cual estará enmarcada la investigación lo constituye el mecanismo para integrar y relacionar datos en Internet.

Para dar solución a la problemática planteada se ha definido como **objetivo general** de la investigación: Proponer un mecanismo de relación de datos haciendo uso de *Linked data* para la integración del sistema de noticias Mella con la Red Social Universitaria en la Universidad de las Ciencias Informáticas.

Del objetivo general anterior se derivan los siguientes **objetivos específicos**:

1. Realizar un estudio conceptual sobre los mecanismos de integración de aplicaciones informáticas con las redes sociales.
2. Proponer un mecanismo de integración entre el sistema de prensa digital Mella de la Universidad de las Ciencias Informáticas con la Red Social Universitaria.
3. Validar la propuesta realizada tomando como referencia los resultados obtenidos por algunas aplicaciones que usan este mecanismo.

La investigación se sostiene en la siguiente **idea a defender**:

La utilización de *Linked data* como mecanismo para la relación de datos permitirá integrar el sistema de noticias Mella con la Red Social Universitaria en la Universidad de las Ciencias Informáticas.

Para estudiar los mecanismos que permiten integrar y relacionar datos en Internet y poder guiar la investigación se utilizaron los siguientes **métodos científicos**:

Métodos Teóricos

- El método **Histórico-lógico** posibilitará hacer un estudio de los fundamentos teóricos de los mecanismos que permiten integrar y relacionar datos en Internet así como las tendencias actuales. Se analizarán las mejores opciones en el momento de seleccionar el componente y se argumentará el por qué de su elección para dar solución a las necesidades de la UCI.
- El **Analítico-Sintético** permitirá extraer y analizar los conceptos y definiciones más importantes relacionadas con el objeto de estudio a partir de los cuales se obtendrá una propuesta que dará cumplimiento a la situación problemática.

Métodos Empíricos

- Mediante la **Entrevista** al cliente se pudo identificar las necesidades reales y se obtuvo la información relacionada con el funcionamiento de los sistemas que se necesitan integrar.

Aportes prácticos esperados del trabajo

Entre los principales beneficios que traerá la presente investigación se encuentran:

- Proporcionará un marco teórico conceptual acerca de la integración de sistemas informáticos con las redes sociales en Cuba y el Mundo.
- Propondrá un estándar para la descripción de las noticias en la Red Social Universitaria.
- Facilitará la comunicación entre el núcleo de la Red Social Universitaria y el sistema de noticias.

Estructura del Documento:

La presente investigación consta de resumen, introducción, tres capítulos, conclusiones, recomendaciones, referencias bibliográficas y bibliografía consultada.

Capítulo I. Fundamentación teórica

Consiste en llevar a cabo la fundamentación teórica del tema a investigar, a partir de un estudio crítico y valorativo, este debe incluir un minucioso estudio del estado del arte a nivel nacional e internacional así como en la Universidad. Teniendo en cuenta las tendencias, tecnologías, herramientas y

programas actuales. Además debe incluir la fundamentación del uso de la tecnología escogida para la realización de ésta investigación.

Capítulo II. Descripción de la propuesta

Se describen las principales funcionalidades que ofrece el sistema de noticias implementado en la Universidad de las Ciencias Informáticas, como punto de partida para enunciar la propuesta de integración con la Red Social Universitaria; exponiendo la forma en que debe llevarse a cabo.

Capítulo III. Validación de la propuesta

Con el fin de avalar la selección del mecanismo de integración como propuesta de solución se realiza un estudio de los resultados obtenidos por algunas aplicaciones que hacen uso del mismo.

Capítulo 1. *Fundamentación Teórica*

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

1.1. INTRODUCCIÓN

En el presente capítulo se pretende abordar los aspectos y conceptos generales del tema a investigar, a partir de un estudio crítico y valorativo, este incluye un estudio del estado del arte, teniendo en cuenta las tendencias actuales. Incluye también la fundamentación del uso de la tecnología escogida para la realización de ésta investigación, dada la factibilidad y ventajas informáticas que la misma ofrece.

1.2. REDES SOCIALES

Las Redes son formas de interacción social, definida como un intercambio dinámico entre personas, grupos e instituciones en contextos de complejidad. Un sistema abierto y en construcción permanente que involucra a conjuntos que se identifican en las mismas necesidades y problemáticas y que se organizan para potenciar sus recursos.

Una sociedad fragmentada en minorías aisladas, discriminadas, que ha desvitalizado sus redes vinculares, con ciudadanos carentes de protagonismo en procesos transformadores, se condena a una democracia restringida. La intervención en red es un intento reflexivo y organizador de esas interacciones e intercambios, donde el sujeto se funda a sí mismo diferenciándose de otros (ZAMORA, 2006).

Las redes sociales en Internet son comunidades virtuales donde sus usuarios interactúan con personas de todo el mundo con quienes encuentran gustos o intereses en común (CASTRO, 2012).

Una red social online es un conjunto de nodos interconectados formando una estructura social de individuos o instituciones organizada culturalmente con un propósito común, que puede ser solidario o no. La red social online despliega y hace posible una forma específica de interacción social y cultural gracias a un número de nodos diverso y reconfigurable -añadiendo o eliminando- de unidades sociales que son los protagonistas de la actividad de la red (FRESNO, 2010).

Luego de analizar las definiciones anteriores se llega a la conclusión de que las redes sociales en Internet están conformadas por grupos de personas u organizaciones dentro de la sociedad, sin tener en cuenta su posición social o profesional, que son capaces de compartir ideas, intercambiar opiniones y enriquecer su cultura informática mediante la utilización de las TIC.

Capítulo 1. Fundamentación Teórica

Tan importante fenómeno mundial le brinda al usuario disímiles ventajas, puesto que permite crear amistades, lograr una mayor difusión de ideas y motiva el debate, favorecen el aprendizaje y contribuyen al desarrollo social, además de que permite mantenerse informados, cuestión importante para el hombre moderno. Tal es la aceptación que han alcanzado las redes sociales, que se utilizan en casi todas las áreas del desarrollo social, según el portal de noticias Ciudad y poder³ las más populares en orden ascendente son:

1. YouTube, la red social más potente que existe, y decimos que es potente porque es la apuesta más grande que hacen las empresas para promocionar sus servicios, grandes personajes como el presidente Obama de Estados Unidos han usado este servicio para promocionar sus servicios obteniendo un índice alto de éxito.
2. Facebook⁴, sitio web de redes sociales donde los usuarios pueden participar en una o más redes sociales, en relación con su situación académica, su lugar de trabajo o región geográfica. Cuenta con más de 700.000.000 usuarios.
3. Twitter⁵, servicio de *microblogging*⁶ que permite enviar mensajes de texto plano de corta longitud, con un máximo de 140 caracteres, llamados *tweets*⁷. Actualmente cuenta con 200.000.000 de usuarios únicos.
4. LinkedIn⁸, sitio web orientado a negocios, en el 2011 se convirtió en la primera red social en poner acciones en la bolsa de valores y cuenta con aproximadamente 100.000.000 usuarios.
5. Hi5⁹, sitio centrado en juegos sociales y, abierto a los desarrolladores de nuevos juegos, cuenta con 11.500.000 usuarios. Quizá el mayor éxito que ha tenido esta red social son sus aplicaciones, sin embargo otros servicios han llegado a opacar esta enorme red social.
6. MySpace¹⁰, de 2005 a 2008, MySpace fue la red social más visitada en el mundo, y en junio del 2006 sobrepasó a Google como el sitio más visitado en la red en los EEUU, tiene un promedio de 80.500.000 usuarios. Una de las redes sociales con más éxito por la capacidad de compartir

³ www.ciudadypoder.com.mx/

⁴ <http://www.facebook.com/>

⁵ <http://www.twitter.com/>

⁶ Servicio que permite a sus usuarios enviar y publicar mensajes breves

⁷ Mensajes cortos enviados por los usuarios en la red social Twitter

⁸ <http://www.linkedin.com/>

⁹ Hi5.com/

¹⁰ <http://www.myspace.com/>

Capítulo 1. Fundamentación Teórica

música, fotos videos y personalizar a fondo nuestro perfil, y que además cuenta con un enorme grupo de desarrollo, actualmente su propietaria es *News Corporation*.

7. Menéame¹¹: noticias al instante, una de las redes de actualidad más famosas que hay. Menéame es un servicio en el cuál cualquier persona puede enviar una noticia y otros usuarios tienen acceso a calificarla y a comentar sobre esta noticia.
8. Google Plus¹², tiene alrededor de 32.000.000 usuarios.
9. Tagged¹³, sitio web de descubrimiento social que cuenta con 25.000.000 usuarios.
10. Orkut¹⁴, red social promovida por Google desde enero del 2004, permite a sus integrantes mantener sus relaciones existentes y hacer nuevos amigos, contactos comerciales o relaciones más íntimas, cuenta con 15.500.000 usuarios.

1.3. SISTEMA DE GESTIÓN DE CONTENIDO (CMS)

Un sistema de gestión de contenidos (*Content Management System*) es un programa que permite crear una estructura de soporte para la creación y administración de contenidos, principalmente en páginas web, por parte de los administradores, editores, participantes y demás roles.

Consiste en una interfaz que controla una o varias bases de datos donde se aloja el contenido del sitio web. El sistema permite manejar de manera independiente el contenido y el diseño.

Tienen 2 formas de clasificarse, por sus características y por su uso y funcionalidad. A nivel mundial existen varios CMS, pero los más usados son Drupal, Joomla y Wordpress, que son precisamente los más empleados en la UCI para la construcción de portales, sitios web y plataformas de creación de blogs.

Entre los contenidos que pueden ser creados y administrados por los CMS, la investigación se centrará en la gestión de noticias como tipo de contenido específico y en el CMS Wordpress como gestor, por ser el utilizado en el desarrollo del sistema de gestión de noticias Mella.

1.3.1. WORDPRESS

¹¹ <http://www.meneame.net>

¹² <https://plus.google.com/>

¹³ <http://www.tagged.com/>

¹⁴ <http://www.orkut.com/>

Capítulo 1. Fundamentación Teórica

Wordpress es un sistema de gestión de contenido enfocado a la creación de blogs (sitios web periódicamente actualizados). Desarrollado en PHP y MySQL, bajo licencia GPL¹⁵ y código modificable. Es un sistema que provee una plataforma fácil de usar para sus publicaciones personales con un enfoque en lo estético, estándares manejados en la web y en la usabilidad.

Aunque especialmente Wordpress se usa para sitios personales y blogs, en los últimos años se comienza a utilizar para portales y sitios de organizaciones que solo desean publicar contenido.

Wordpress se ha convertido en el CMS más popular de la blogosfera¹⁶ y en el más popular con respecto a cualquier otro CMS de aplicación general. Las causas de su enorme crecimiento son, entre otras, la licencia, facilidad de uso, características como gestor de contenidos, la utilización que se le da en la creación de plataformas de blog y medios de prensa y porque representa una avanzada plataforma semántica de publicación personal orientada a la estética, los estándares web y la usabilidad (AMADOR, 2008).

1.3.2. TECNOLOGÍAS DE DESARROLLO

Lenguajes a utilizar

Los lenguajes de programación son indispensables para el desarrollo de programas informáticos, pues sin ellos no existirían los diferentes programas (softwares) que requieren las computadoras, de ahí que su selección es importante antes del desarrollo de cualquier sistema. Como la funcionalidad será implementada sobre Wordpress, los lenguajes que se utilizarán son los propiamente definidos para el desarrollo de este CMS.

Lenguaje de Marcado de Hipertexto (HTML) 4.0

HTML (*Hyper Text Markup Language*) es el lenguaje con el que se escriben las páginas web. Estas pueden ser vistas por el usuario mediante un tipo de aplicación llamada navegador. El HTML es el lenguaje usado por los navegadores para mostrar las páginas web al usuario, siendo hoy en día la interface más extendida en la red.

¹⁵ GPL o Licencia Pública general (*General Public License*) es una licencia que permite modificar el programa para adaptarlo a las propias necesidades de los usuarios ya que este viene acompañado de su código fuente.

¹⁶ Blogosfera: un sistema virtual, en el que se establecen comunidades de weblogs, categorizados temáticamente o por perfiles de interés

Capítulo 1. Fundamentación Teórica

Este lenguaje permite aglutinar textos, sonidos e imágenes y combinarlos a gusto. Además, y es aquí donde reside su ventaja con respecto a libros o revistas, el HTML permite la introducción de referencias a otras páginas por medio de los enlaces hipertexto (ÁLVAREZ, 2001).

JavaScript 1.5

JavaScript es un lenguaje de programación utilizado para crear pequeños programas encargados de realizar acciones dentro del ámbito de una página web.

Se trata de un lenguaje de programación del lado del cliente, porque es el navegador el que soporta la carga de procesamiento. Gracias a su compatibilidad con la mayoría de los navegadores modernos, es el lenguaje de programación del lado del cliente más utilizado.

Con JavaScript podemos crear efectos especiales en las páginas y definir interactividades con el usuario. El navegador del cliente es el encargado de interpretar las instrucciones JavaScript y ejecutarlas para realizar estos efectos e interactividades, de modo que el mayor recurso, y tal vez el único, con que cuenta este lenguaje es el propio navegador.

Entre las acciones típicas que se pueden realizar en JavaScript tenemos dos vertientes. Por un lado los efectos especiales sobre páginas web, para crear contenidos dinámicos y elementos de la página que tengan movimiento, cambien de color o cualquier otro dinamismo. Por el otro, JavaScript permite ejecutar instrucciones como respuesta a las acciones del usuario, con lo que se puede crear páginas interactivas con programas como calculadoras, agendas, o tablas de cálculo (ÁLVAREZ, 2001).

XML

XML (Lenguaje de Marcas Extensible). Se trata de un metalenguaje (un lenguaje que se utiliza para decir algo sobre otro lenguaje) extensible de etiquetas que fue desarrollado por el *World Wide Web Consortium*¹⁷ (W3C), un consorcio internacional¹⁸ que elabora recomendaciones para la *World Wide Web*¹⁸ (www).

El XML es una adaptación del SGML (*Standard Generalized Markup Language*), un lenguaje que permite la organización y el etiquetado de documentos. Esto quiere decir que el XML no es un lenguaje en sí mismo, sino un sistema que permite definir lenguajes de acuerdo a las necesidades. El XHTML, el MathML y el SVG son algunos de los lenguajes que el XML está en condiciones de definir.

¹⁷ W3C: Consorcio Internacional de la Red Informática Mundial.

¹⁸ www: Red Informática Mundial.

Capítulo 1. Fundamentación Teórica

Las bases de datos, los documentos de texto, las hojas de cálculo y las páginas web son algunos de los campos de aplicación del XML. El metalenguaje aparece como un estándar que estructura el intercambio de información entre las diferentes plataformas (*Definición de XML*, 2008).

Hojas de Estilo en Cascada (CSS) 3.0

Hojas de Estilo en Cascada (*Cascading Style Sheets*), es un mecanismo simple que describe cómo se va a mostrar un documento en la pantalla, o cómo se va a imprimir, o incluso cómo va a ser pronunciada la información presente en ese documento a través de un dispositivo de lectura. Esta forma de descripción de estilos ofrece a los desarrolladores el control total sobre estilo y formato de sus documentos.

CSS se utiliza para dar estilo a documentos HTML y XML. Los estilos definen la forma de mostrar los elementos HTML y XML. CSS permite a los desarrolladores web controlar el estilo y el formato de múltiples páginas web al mismo tiempo. Cualquier cambio en el estilo marcado para un elemento en la clase CSS afectará a todas las páginas donde aparezca un elemento vinculado a esa clase CSS (*Guía breve de CSS*, 2008).

PHP (*Hypertext pre-processor*) 5.3.10

PHP acrónimo recursivo que significa *Hypertext Pre-processor*, es un lenguaje de programación usado normalmente para la creación de contenido para sitios web con los cuales se puede programar las páginas HTML y los códigos de fuente. Se trata de un lenguaje interpretado usado para la creación de aplicaciones para servidores, o creación de contenido dinámico para sitios web. Últimamente también para la creación de otro tipo de programas incluyendo aplicaciones con interfaz gráfica usando las librerías Qt13 o GTK+14.

Sus principales características son:

- Es un lenguaje multiplataforma.
- Capacidad de conexión con la mayoría de los manejadores de base de datos que se utilizan en la actualidad, destaca su conectividad con MySQL.
- Leer y manipular datos desde diversas fuentes, incluyendo datos que pueden ingresar los usuarios desde formularios HTML.

Capítulo 1. Fundamentación Teórica

- Capacidad de expandir su potencial utilizando la enorme cantidad de módulos (llamados ext's o extensiones).
- Posee una amplia documentación en su página oficial, entre la cual se destaca que todas las funciones del sistema están explicadas y ejemplificadas en un único archivo de ayuda.
- Es libre, por lo que se presenta como una alternativa de fácil acceso para todos.
- Permite las técnicas de Programación Orientada a Objetos.
- Permite crear formularios para la web.
- Biblioteca nativa de funciones sumamente amplia e incluida.
- No requiere definición de tipos de variables, ni manejo detallado del bajo nivel (*PHP (PHP Hypertext Pre-processor)*, 2002).

1.3.3. HERRAMIENTAS DE DESARROLLO

Servidor web

Un servidor web o servidor HTTP, es un *software* que se encuentra en ejecución constantemente en un ordenador, el cual generalmente utiliza el protocolo de comunicaciones HTTP (*Hypertext Transference Protocol*) de la capa de aplicación del Modelo OSI¹⁹ para recibir peticiones de información de un *software* cliente (navegador) en el ordenador del usuario. El servidor web se encarga de responder a estas peticiones de forma adecuada, entregando como resultado una página web o información de todo tipo de acuerdo a los comandos solicitados.

Apache 2.2.14

El Servidor Apache HTTP es un servidor web de tecnología *Open Source* sólido y para uso comercial desarrollado por la *Apache Software Foundation* en el proyecto HTTP Server. Apache es usado principalmente para enviar páginas web estáticas y dinámicas a la *World Wide Web* y es el servidor web más usado en todo el mundo. A continuación se exponen las principales características de este *software*:

¹⁹ Modelo de interconexión de sistemas abiertos. Es un marco de referencia para la definición de arquitecturas de interconexión de sistemas de comunicaciones.

Capítulo 1. Fundamentación Teórica

- Funciona en una multitud de Sistemas Operativos, lo que lo hace prácticamente universal.
- Apache es una tecnología gratuita de código abierto. El hecho de ser gratuita es importante pero no tanto como que se trate de código fuente abierto. Esto le da una transparencia a este *software* de manera que si queremos ver que es lo que estamos instalando como servidor, lo podemos saber, sin ningún secreto.
- Apache es un servidor altamente configurable de diseño modular. Es muy sencillo ampliar las capacidades del servidor web Apache. Actualmente existen muchos módulos para Apache que son adaptables a este. Otra cosa importante es que cualquiera que posea una experiencia modesta en la programación de C o Perl puede escribir un módulo para realizar una función determinada.
- Apache te permite personalizar la respuesta ante los posibles errores que se puedan dar en el servidor. Es posible configurar Apache para que ejecute un determinado script cuando ocurra un error en concreto.
- Tiene una alta configurabilidad en la creación y gestión de logs²⁰. Apache permite la creación de ficheros de log a medida del administrador, de este modo puedes tener un mayor control sobre lo que sucede en tu servidor (CIBERAULA, 2010).

Netbeans (IDE) 7.0.1

Netbeans es un entorno de desarrollo o una herramienta para que los programadores puedan escribir, compilar, depurar y ejecutar programas. Está escrito en Java pero puede servir para cualquier otro lenguaje de programación. Existe además un número importante de módulos para extender el Netbeans. Es un producto libre y gratuito sin restricciones de uso (NETBEANS).

Principales características de Netbeans:

- Creación de Proyectos PHP: Integra lenguajes como HTML, JavaScript y CSS.
- Integración con Symfony y ZendFramework: Agiliza el proceso de desarrollo de aplicaciones con estos populares Framework de PHP. Gracias a Netbeans ya es posible dejar la consola de

²⁰ Un **log** es un registro oficial de eventos durante un rango de tiempo en particular. Para los profesionales en seguridad informática es usado para registrar datos o información sobre quién, qué, cuándo, dónde y por qué un evento ocurre para un dispositivo en particular o aplicación.

Capítulo 1. Fundamentación Teórica

comandos de Symfony y centrarse en desarrollar en el IDE, además de que se encuentran cargadas todas las clases, ayuda en línea, etc.

- Editor de Código Fuente: El editor de PHP es ágil y a la vez robusto, contiene ayuda en línea, reconocimiento de sintaxis y todo lo que proveen las últimas versiones de PHP.
- Integración con PHP Unit Testing: Es posible crear test con *PHPUnit*²¹, para diferentes funciones, luego realizar la comprobación y ver todos los resultados.
- Depuración de PHP: Netbeans integra muy bien la utilización *Xdebug*²², gracias a esto podemos inspeccionar y examinar cada variable local, establecer puntos de interrupción y evaluar el código en nuestra lógica. El IDE de Netbeans para PHP también ofrece la línea de comandos de depuración, la salida del programa PHP aparece en una pantalla de línea de comandos en el IDE de sí mismo y se puede inspeccionar el código HTML generado sin tener que cambiar a un navegador.
- Integración con MySQL: Netbeans posee una integración completa en términos de administración básica y avanzada de MySQL, y todo desde el mismo entorno (MALDONADO, DANIEL, 2010).

1.3.4. SISTEMAS GESTORES DE BASE DE DATOS

Sistema Gestor de Bases de Datos (SGBD) o DBMS (*Data Base Management System*) es un conjunto de programas que permiten crear y mantener una base de datos, asegurando la integridad, confidencialidad y seguridad de los datos. Está compuesto por un lenguaje de definición de datos, un lenguaje de manipulación de datos y un lenguaje de consulta y su objetivo principal es servir de interfaz entre la base de datos, el usuario y las aplicaciones (IBARRA, 2010).

Se selecciona MySQL en su versión 5.0.6 debido a que el CMS Wordpress previamente definido para la solución, utiliza solamente este Sistema Gestor de Base de Datos.

²¹ **PHPUnit** es un entorno para realizar pruebas unitarias en el lenguaje de programación PHP. Es un framework de la familia xUnit originada con SUnit de Kent Beck.

²² **Xdebug** es una extensión PHP que provee capacidades de depuración y perfilado utilizando el protocolo de depuración DBGp.

Capítulo 1. Fundamentación Teórica

MySQL 5.0.6

MySQL es un sistema de gestión de base de datos relacional, que cuenta con licencia *GPL* de la *GNU*²³.

Este gestor de bases de datos es, probablemente, el gestor más usado en el mundo del *software* libre, debido a su gran rapidez y facilidad de uso. Esta gran aceptación es debida, en parte, a que existen infinidad de librerías y otras herramientas que permiten su uso a través de gran cantidad de lenguajes de programación, además de su fácil instalación y configuración (ACOSTA, 2008).

1.4. WEB SEMÁNTICA

La Web Semántica se está convirtiendo en un paradigma que viene a romper la curva de la experiencia actual para desarrollar el potencial de Internet como fuente de conocimiento.

Es un conjunto de actividades desarrolladas en el contexto de *W3C*, el concepto fue creado por Tim Berners-Lee creador de la *www*, este la define a grandes rasgos como un proyecto que parte de la idea de que las páginas web destinadas a los humanos, podrían también ser datos entendibles por las máquinas.

Con la ayuda de la Web Semántica se han desarrollado un conjunto de tecnologías que se aplican en muchos ámbitos: redes sociales, publicación de datos, realización de inferencias, marcado semántico de documentos convencionales. La misma se basa en la idea de añadir metadatos semánticos y ontológicos a la *www*.

El objetivo de la Web Semántica es mejorar Internet ampliando la interoperabilidad entre los sistemas informáticos usando "agentes inteligentes". Agentes inteligentes son programas en las computadoras que buscan información sin operadores humanos.

La Web Semántica representa una web con mayor significado que posibilita que cualquier usuario en Internet pueda encontrar respuestas sencillas y rápidas a cualquier interrogante o problema que se le presente. Al dotar a la Web de más significado y por tanto de más semántica, se obtienen soluciones a problemas habituales en la búsqueda de la información, gracias a un proceso común que permite que se pueda compartir, procesar y transferir información de forma más sencilla (CODINA, 2009).

²³ GNU es un proyecto iniciado por Richard Stallman con el objetivo de crear un sistema operativo completamente libre.

Capítulo 1. Fundamentación Teórica

Es por esta razón que se llega a la conclusión que el objetivo principal de la Web Semántica es el de desarrollar lenguajes de expresión para describir información en formas entendibles para las PC.

En estos momentos en la Web existen 2 grandes problemas, uno es la sobrecarga de información existente y el otro es la diversidad de fuentes de información, los cuales la Web Semántica ayuda a resolver, pues dispone de tecnologías de descripción de los contenidos, como RDF (*Resource Description Framework*²⁴) y OWL (*Web Ontology Language*²⁵), además de XML, el lenguaje de marcas diseñado para describir los datos. Estas tecnologías se combinan para aportar descripciones explícitas de los recursos de la Web (ya sean estos catálogos, formularios, mapas u otro tipo de objeto documental).

En el 2001 al popularizar el término mundialmente, Tim Berners-Lee propone que la arquitectura de la Web Semántica podría representarse en capas:

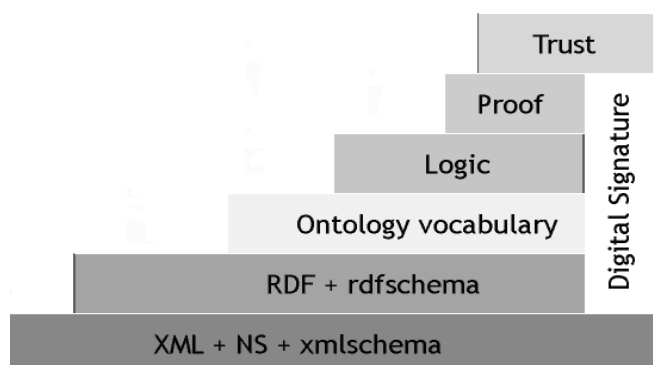


Figura 1. Arquitectura de la Web Semántica.

- Capa XML: se encuentra en el nivel inferior y es la que permitirá estructurar los datos.
- Capa RDF: es la capa siguiente, tiene a su cargo la definición de la semántica de los datos.
- Capa Ontológica: define las relaciones entre los recursos.
- Capa Lógica: es la capa superior y define las reglas y mecanismos para hacer inferencias sobre los datos.

²⁴ RDF: Marco de Descripción de Recursos

²⁵ OWL: Lenguaje Ontológico para la Web

Capítulo 1. *Fundamentación Teórica*

- Capas Prueba y Veracidad: ejecutan y evalúan juntas las reglas de la capa lógica para determinar la fiabilidad del recurso. Aquí es donde interviene la firma digital, el concepto de comercio seguro.
- Firma digital: bloque encriptado de datos que serán utilizados por los ordenadores y los agentes para verificar que la información adjunta ha sido ofrecida por una fuente específica confiable (LAPUENTE, 2011).

Como se puede observar en esta arquitectura los pilares que sostienen a la Web Semántica son las ontologías y las anotaciones. Las ontologías, que normalmente son representadas en el esquema RDF o lenguaje OWL, describen formalmente conceptualizaciones compartidas de un dominio (por ejemplo, personas, reuniones, etc.) Las anotaciones por otra parte permiten la unión en las descripciones basadas en ontologías a los actuales recursos de la Web (por ejemplo, diciendo que la información contenida en una página web se refiere a una persona) y se describen normalmente en RDF.

Capítulo 1. Fundamentación Teórica

1.4.1. MAPA CONCEPTUAL DE LA WEB SEMÁNTICA



Figura 2. Mapa conceptual de la Web Semántica

Capítulo 1. Fundamentación Teórica

1.5. ONTOLOGÍAS

Las ontologías están consideradas como los pilares de la Web Semántica, aunque universalmente no tengan una definición aceptada.

En la Web Semántica un vocabulario es considerado como una forma especial de ontología.

Según Gruber y Studer la ontología es: *“una especificación explícita y formal sobre una conceptualización compartida”* es decir, que las ontologías se encargan de definir conceptos y relaciones de un dominio de forma compartida, y que esta conceptualización debe ser presentada de una forma legible y utilizable por los ordenadores.

En inteligencia artificial, donde además surgió el término, una ontología es una descripción de los conceptos y relaciones que existen para un agente o una comunidad de agentes, así su función principal consiste en facilitar el intercambio de información entre los agentes.

Gracias a las ontologías y las relaciones que establecen entre distintos componentes, las computadoras serán capaces de comprender el significado de la semántica asociada a una web siguiendo los vínculos especificados por las mismas. Son capaces de definir información sobre equivalencias semánticas, de esta manera si se realizara una búsqueda en Internet de la palabra "computadora" los resultados incorporarían además los resultados arrojados por términos equivalentes como "ordenador" o "computador personal". En el fondo lo que se pretende es lograr una red de información con un pensamiento abstracto lo más parecido posible al del ser humano, es decir, que el sistema, dados ciertos términos para la búsqueda devuelva información relacionada que resulte relevante de acuerdo a los objetivos que se plantearon.

Típicamente, las ontologías en las computadoras se relacionan estrechamente con vocabularios fijos, con cuyos términos debe ser descrito todo lo demás. Debido a que esto puede ocasionar representaciones pobres para ciertos dominios de problemas, se deben crear esquemas más especializados para convertir en útiles los datos a la hora de tomar decisiones en el mundo real.

Las mismas tienen algunos componentes que ayudarán a entender y representar el conocimiento de algún dominio, por ejemplo:

- **Conceptos:** son ideas básicas que se intentan formalizar. Pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento.

Capítulo 1. Fundamentación Teórica

- Relaciones: representan la interacción y enlace entre los conceptos del dominio. Suelen formar la taxonomía²⁶.
- Funciones: son un tipo de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología.
- Instancias: se utilizan para representar objetos determinados de un concepto.
- Axiomas: son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: “Si A y B son de la clase C, entonces A no es subclase de B”, “Para todo A que cumpla la condición CL, A es B”. También permiten junto a la herencia de conceptos, inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos (TELLO, 2001).

1.5.1. LENGUAJES DE ONTOLOGÍAS

Las ontologías requieren de un lenguaje lógico y formal para ser expresadas. En inteligencia artificial se han desarrollado numerosos lenguajes para este fin, algunos basados en la lógica de predicados, otros basados en *frames* (marcos), que tienen un mayor poder expresivo, pero menor poder de inferencia; e incluso existen lenguajes orientados al razonamiento. Todos estos lenguajes han servido para desarrollar otros lenguajes aplicables a la web. A continuación se mencionan algunos de ellos:

- RDF
- OWL
- DAML²⁷ (*DARPA Agent Markup Language*)
- OIL²⁸ (*Ontology Inference Layer*)

En estos últimos años se han desarrollado múltiples lenguajes y estándares para la definición de las ontologías, y parece ser que el estándar preferido es el RDF.

RDF es un esquema para definir e intercambiar metadatos basado en XML, que se ocupa de la semántica y permite, mediante relaciones de clasificación, crear una jerarquía de conceptos de un

²⁶ Es, en su sentido más general, la ciencia de la clasificación.

²⁷ Lenguaje agente de marcas de la DARPA.

²⁸ Capa de inferencia ontológica.

Capítulo 1. Fundamentación Teórica

dominio. En pocas palabras, es el modelo que establece el formato de datos que permite al usuario crear lenguajes entendibles para las computadoras de manera que estas puedan procesar y clasificar semánticamente la información. Como su nombre lo indica (Marco de descripción de recursos) su principal objetivo es la descripción de los recursos en Internet, es decir, una imagen, un documento, una animación, una película, un autor, que pueden ser descritas en término de propiedades y reconocibles particularmente a través del URI²⁹.

RDF se puede clasificar en diferentes formatos:

- RDF/XML: sintaxis XML para representar RDF.
- RDFa: conjunto de extensiones de XHTML propuestas por el W3C para introducir semántica en los documentos.
- N3: serialización no-XML de modelos RDF.
- Turtle: formato de serialización para grafos RDF.
- JSON: formato ligero para el intercambio de datos.

Todo el sistema RDF parte de tres entidades lógicas:

- **Recursos:** todas las cosas descritas por expresiones RDF se denominan recursos. Los recursos representan cualquier entidad (lugares, personas, objetos) del mundo real y están identificados por un URI. Un recurso puede ser una página web, una parte de una página web o un objeto que no sea accesible vía web, por ejemplo un libro impreso.
- **Propiedades:** una propiedad es un aspecto específico, característica, atributo, o relación utilizado para describir un recurso. Cada propiedad tiene un significado específico, define sus valores permitidos, los tipos de recursos que puede describir, y sus relaciones con otras propiedades.
- **Sentencias:** un recurso específico junto con una propiedad denominada, más el valor de dicha propiedad para ese recurso es una sentencia RDF. Estas tres partes individuales de una sentencia se denominan, respectivamente, sujeto, predicado y objeto. Una propiedad es a su

²⁹ URI: las referencias. URI es el acrónimo de "Uniform Resource Identifier" o Identificador Uniforme de Recursos, identificador único que permite la localización de un recurso que puede ser accedido vía Internet. Se trata del URL (descripción de la ubicación) más el URN (descripción del espacio de nombre).

Capítulo 1. Fundamentación Teórica

vez un recurso. El objeto es el valor asignado a dicha propiedad, y puede ser una cadena simple de caracteres, u otro recurso, es decir, un recurso especificado por un URI.

Las sentencias constituyen la construcción básica que establece el modelo RDF. Es decir, el significado de los datos se expresa mediante un conjunto de sentencias RDF que formablemente son representadas por tripletas (sujeto, predicado, objeto) (RUIZ, 2008).

1.5.2. ESTÁNDARES DE METADATOS DISPONIBLES (WEB SEMÁNTICA)

- Elementos de datos: DC, UDEF, PRISM.
- Estándares XML: NITF, NewsML, RIXML, XBRL, ebXML, IMS (PABLO CASTELLS, 2003).

1.5.3. ESTÁNDARES XML

La Web Semántica propone disímiles estándares con base en el formato XML, para representar la estructura y el formato de contenido de los artículos y documentos noticiosos. Estos estándares nos brindan una gran ventaja pues cumplen con casi todas las demandas que un editor se pueda imaginar, para un estándar que sirva para el intercambio de todo tipo de información de interés periodístico.

➤ **NITF (News Industry Text Format)**

El formato de texto para la industria noticiosa (NITF) está basado en XML, desarrollado por el Consejo Internacional de Prensa y Telecomunicaciones (IPTC), una asociación internacional independiente de agencias de noticias y editoriales. Estándar abierto que representa el contenido y la estructura de los artículos de noticias y los metadatos de acompañamiento.

La intención de NITF es dar un formato de base a las noticias, el cual puede ser cambiado por los editores, los que pueden adaptar la apariencia, sensación, la interactividad y los dispositivos a la banda ancha, cumpliendo con las necesidades de sus abonados.

Los documentos NITF se pueden traducir en HTML, WML (para dispositivos inalámbricos), RTF (para imprimir), o cualquier otro formato que el editor desee.

Es el formato preferido para los artículos de noticias individuales, para la retroalimentación de noticias. Para su utilización no necesita de una licencia o de patentes. No depende de ningún otro estándar. Incluye soporte HTML compatible con tablas y listas.

Capítulo 1. *Fundamentación Teórica*

Soporta conexiones a los objetos de medios externos, tales como imágenes, audio o vídeo. También admite la incrustación de objetos multimedia (PRESERVATION, 2004).

➤ **XMLNews**

XMLNews es un conjunto de especificaciones para el intercambio de objetos de comunicación, tales como historias, imágenes o clips de audio en un formato estándar a través de diferentes aplicaciones y sistemas operativos. XMLNews utiliza XML y consta de dos partes: XMLNews-Story y XMLNews-Meta.

XMLNews-Story es un tipo de documento XML para texto basado en noticias e información. Se define el formato del contenido de una noticia y es un subconjunto del formato NITF, este documento XML, está diseñado para marcar y entregar el contenido de noticias en una variedad de formas, incluyendo impresión, dispositivos inalámbricos, y la Web. XMLNews-Meta define el formato de los metadatos asociados con una historia (o cualquier otro tipo de objeto de noticias) y se basa en RDF.

Las noticias pueden llegar de muchas formas diferentes, incluyendo texto, imágenes, sonido, video, mundos virtuales en tres dimensiones, guiones interactivos, o casi cualquier otra cosa que los proveedores de noticias se puedan imaginar (ROUSE, 2005).

➤ **NewsML**

NewsML es una palabra inventada, que combina "Noticias" y "XML". Como un formato XML especial de transmisión estándar, NewsML está especializado en la aplicación de intercambio de noticias. NewsML es para entregar el contenido automáticamente y como una multimedia de los proveedores de contenido a los clientes. El mismo tiene sus propios campos de descripción para el origen, fecha, autor, título y otras clasificaciones de la noticia. NewsML fue aprobado por IPTC en octubre de 2000. En el grupo de trabajo de IPTC están casi todos los proveedores de contenido importantes, sobre todo las agencias de noticias alemanas e internacionales.

Esencial para NewsML, XML es el salto cualitativo hacia HTML. Considerando que el tradicional HTML sólo muestra contenidos, XML estructura los contenidos, similar a una base de datos. Con XML, el lenguaje de descripción de una página se completa con la estructuración de los datos integrados.

Sin embargo, los datos se estructuran de manera diferente para cada tarea y cada cliente y por lo tanto se estructuran diferente en XML. El formato NewsML estandarizado está organizado de tal forma que cumple casi todas las demandas imaginables en un estándar para el intercambio de noticias. En contraste con el estándar NIFT, NewsML es especialmente preparado para la era de la multimedia.

Capítulo 1. Fundamentación Teórica

Este estándar es una estructura base, dejando el diseño y otros elementos de presentación a hojas de estilo y otros idiomas de presentación, tales como XHTML o SMIL (IPTC).

➤ Prism

PRISM por sus siglas en inglés *Publishing Requirements for Industry Standard Metadata* está basado en RDF. Realizado por un consorcio industrial de empresas editoriales y de fabricantes de *software* de publicación entre las que se incluyen Adobe, Quark, Condé Nast y Time Inc. Es un estándar XML parecido a NewsML de un vocabulario de metadatos para la industria editorial que facilita la agregación y sindicación de contenido digital.

Define un estándar para la descripción de contenidos interoperables, el intercambio y la reutilización en los contextos de la edición tradicional y electrónica. PRISM recomienda el uso de ciertos estándares existentes, tales como XML, RDF, el Dublin Core, y varias especificaciones de la ISO para definir lugares, idiomas y formatos de fecha / hora. Más allá de las recomendaciones, define un pequeño número de espacios de nombres XML y vocabularios de valores controlados, con el fin de alcanzar los objetivos mencionados anteriormente (INTERNATIONAL DIGITAL ENTERPRISE ALLIANCE, 2004).

Luego de la explicación de los cuatro estándares más utilizados mundialmente para el intercambio de noticias, se llega a la conclusión de que el más idóneo para cumplimentar las necesidades de la integración entre los sistemas es el NewsML, por ser experimentado en la exportación e intercambio de noticias, trabajar de conjunto con los otros estándares, definir sus propios campos de descripción de la noticia y por tener la capacidad de exportar los contenidos en cualquier formato.

1.6. MECANISMOS DE INTEGRACIÓN ENTRE APLICACIONES

1.6.1. RSS

RSS es un formato XML para syndicar o compartir contenido en la web. Se utiliza para difundir información actualizada frecuentemente a usuarios que se han suscrito a la fuente de contenidos. El formato permite distribuir contenidos sin necesidad de un navegador, utilizando un *software* diseñado para leer estos contenidos. A pesar de eso, es posible utilizar el mismo navegador para ver los contenidos RSS. Las últimas versiones de los principales navegadores permiten leer los RSS sin necesidad de *software* adicional. RSS es parte de la familia de los formatos XML desarrollado específicamente para todo tipo de sitios que se actualicen con frecuencia y por medio del cual se

Capítulo 1. Fundamentación Teórica

puede compartir la información y usarla en otros sitios web o programas. Facilita la gestión y publicación de información y noticias web.

Un RSS es realmente un formato de archivo, basado en XML, que sirve para recoger contenidos publicados en páginas web. Los RSS tienen extensión .rss o bien .xml, pero en realidad son un simple archivo de texto donde aparecen referencias a contenidos publicados, en un formato específico, creado a partir de XML. En el archivo RSS simplemente están los datos de las novedades del sitio, como el título, fecha de publicación o la descripción. El programa que lea el RSS será encargado de darle estilo o apariencia a los datos que se incluyan en el archivo y presentarlos de una manera atractiva al usuario y de fácil lectura.

1.6.2. API (APPLICATION PROGRAMMING INTERFACE)

Una API (*Application Programming Interface* o *Interfase de Programación de Aplicaciones*) es un conjunto de funciones que facilitan el intercambio de mensajes o datos entre dos aplicaciones. Es una forma en la cual dos aplicaciones que trabajan al mismo tiempo se comunican e intercambian datos. Se trata del conjunto de llamadas a ciertas bibliotecas que ofrecen acceso a ciertos servicios desde los procesos y representa un método para conseguir abstracción en la programación. La API era un concepto que se usaba generalmente en sistemas operativos para que los programadores de las aplicaciones y *software* pudieran llamar de forma fácil a componentes ya programados del sistema operativo. Pero donde han tenido y están teniendo un gran éxito actualmente es en las aplicaciones web (BAIL, 2009).

Uno de los servicios web que mejor uso ha hecho de su API y le ha dado un gran crecimiento de programas alternativos para usarlo y de servicios complementarios a su alrededor ha sido Twitter.

➤ **Ventajas de las API**

- Permite que un sitio web brinde determinado servicio a otro.
- Facilita el “*mash-up*”³⁰ o mezcla de servicios.
- Pueden desarrollarse para cualquier plataforma y sistema operativo.

³⁰ página web o aplicación que usa y combina datos, presentaciones y funcionalidad procedentes de una o más fuentes para crear nuevos servicios.

Capítulo 1. Fundamentación Teórica

- Una buena API hace más fácil el trabajo de desarrollo de un programa, ya que provee todos los bloques para construirlo. El programador lo único que hace es poner todos los bloques juntos.
- API está diseñado especialmente para los programadores, ya que garantiza que todos los programas que utilizan API, tendrán interfaces similares. Asimismo, esto le facilita al usuario aprender la lógica de nuevos programas.
- Cuando se realiza una requisición, el servidor llamará al API, brindando la ventaja de disponer de una mayor cantidad de servicios (TROLLTECH, 2008).

1.6.3. LINKED DATA (DATOS VINCULADOS)

Linked data se puede definir como una serie de buenas prácticas para publicar, compartir y conectar conjuntos de datos, información y conocimiento en la Web Semántica utilizando URIs y RDF.

En informática, *Linked data* describe un método para la publicación de datos estructurados, para que de esta manera puedan ser vinculados entre sí, y ser más útiles. Se basa en tecnología web estándar, como las URIs y HTTP, pero en lugar de utilizarlo para páginas web leídas por humanos, se extiende para compartir información en una forma que puede ser leída automáticamente por las computadoras (BERNERS-LEE, 2006).

El objetivo de *Linked data* es el de permitir a las personas compartir los datos estructurados en la web tan fácilmente como se comparten los documentos hoy día en la web.

El término se refiere al estilo de publicación y el mecanismo de interconexión de datos estructurados en la Web.

➤ **Ventajas de *Linked data***

- Facilita el uso, exploración e integración de datos heterogéneos.
- Posibilita el fácil intercambio de datos.
- Con su aplicación se pueden descubrir más datos relacionados durante la consulta de los datos originales y se puede aprender a partir de los esquemas de los datos.

Capítulo 1. Fundamentación Teórica

- Los datos son fácilmente localizables (porque pueden ser vinculadas por otros datos), incrementando así el valor de los datos publicados.

1.7. DIFERENCIAS ENTRE LAS WEB API Y LINKED DATA

Algunas de las importantes fuentes de datos como Google, Yahoo, eBay, ofrecen el acceso a sus datos mediante API. A las API se pueden acceder mediante una gama de diferentes mecanismos, y los datos que se obtienen, se representan a través de distintos formatos de contenido. Contrario a esto, *Linked data* solo trabaja con un pequeño grupo de tecnologías estandarizadas: URIs y HTTP como mecanismo de identificación y acceso y RDF como formato de contenido.

La utilización de un simple conjunto de tecnologías, en lugar de confiar en diversas interfaces y formatos, permite que las fuentes de datos sean fácilmente rastreadas por los motores de búsqueda y de fácil acceso para los navegadores de datos genéricos.

Además de estos detalles técnicos, también existe una gran diferencia contextual entre las API y *Linked data*, pues la mayoría de las API no asignan un único identificador global a los elementos de datos como lo hace *Linked data*, lo que dificulta que se establezcan vínculos entre elementos pertenecientes a diferentes fuentes de datos, con el fin de conectar la información en un espacio de datos global.

En cambio con *Linked data* se pueden descubrir nuevas fuentes de información siguiendo los enlaces RDF tal como aparecen en la web sin necesidad de cambiarle el código a la aplicación.

Por tanto, se llega a la conclusión de que las tecnologías *Linked data* contribuyen a la conexión entre los diferentes conjuntos de datos que existen actualmente en la web, convirtiéndolos en un espacio de información único y global.

Capítulo 2. Propuesta de solución

CAPITULO 2: PROPUESTA DE SOLUCIÓN

2.1. INTRODUCCIÓN

En el presente capítulo se realiza el diagnóstico del campo de acción de la investigación y la descripción de la propuesta de solución. De la misma forma se describen los elementos que brindarán mayor entendimiento de las características y conceptos relacionados con la propuesta para darle cumplimiento a los objetivos trazados.

2.2. DIAGNÓSTICO DEL CAMPO DE ACCIÓN

En la Universidad de las Ciencias Informáticas, para darle cumplimiento al programa de informatización de los principales procesos de la institución, el grupo de trabajo del Centro de Informatización desarrolla como parte de su trabajo, un grupo de aplicaciones y servicios. Entre las temáticas en las que se desenvuelve el centro se encuentra la línea de redes sociales. Esta línea se encarga del desarrollo y mantenimiento de servicios con enfoque de redes sociales que representan componentes de la RSU, y de la misma forma hacer una transformación en los servicios existentes, fortaleciendo su utilización y práctica.

Entre los nuevos servicios creados en la UCI se encuentra el periódico digital Mella cuyo objetivo es el de centralizar de forma organizada la gestión de la información noticiosa.

Dentro de los servicios de redes sociales que brinda esta aplicación se encuentran: dar la posibilidad al usuario de comentar y compartir una noticia, artículo, documento o aplicación, administrar un blog, votar en encuestas, servicio de suscripción, entre otros servicios que conforman lo que se podría caracterizar como espacios de Web Social.

Actualmente los servicios que brinda este sistema actúan de forma independiente en el sitio, sin lograr interactuar con otro sistema, a pesar de constituir uno de los componentes de la todavía en construcción RSU. Constituye esta la limitación que presenta dicha aplicación; al no existir un mecanismo capaz de integrarla con la Red Social Universitaria, no provee ningún método para que otras aplicaciones sean capaces de recuperar la información expuesta; por lo que la información se encuentra duplicada en la red. Por ello en lugar de publicar la misma información varias veces, sería más útil reutilizarla.

Capítulo 2. Propuesta de solución

Esta ha sido a lo largo de todo el proceso de implementación del sistema, una de las preocupaciones que ha tenido el equipo de desarrollo, cómo lograr un mecanismo que posibilite la unión de la Red Social Universitaria y el sistema de noticias Mella.

Por todos los problemas anteriormente tratados y la estrategia de integración basada en la Web Semántica que propone el núcleo de la RSU, es que se opta por el uso de *Linked data* para la vinculación de los contenidos de la RSU y el sistema de noticias Mella.

2.3. DESCRIPCIÓN DE LA PROPUESTA

A partir de los resultados del estudio desarrollado se propone realizar la integración de las aplicaciones y servicios de la RSU mediante la utilización del método *Linked data* y el estándar de intercambio de noticias NewsML.

2.3.1. LINKED DATA

Tim Berners-Lee definió cuatro principios que deben caracterizar la información en *Linked data*:

1. Utilizar URIs para identificar los recursos.
2. Aprovechar el HTTP de la URI para que la gente pueda localizar y consultar los recursos.
3. Proporcionar información útil acerca del recurso cuando se consulte la URI mediante los estándares.
4. Incluir enlaces a otras URIs, de forma que se potencie el descubrimiento de información en la Web (TIM BERNERS-LEE).

Para aplicar estos principios, se utilizan una serie de componentes que aparecen en la definición de los principios:

1. Uniform Resource Identifier (URI): cadena de caracteres corta que identifica unívocamente un recurso.
2. Resource Description Framework (RDF): es un *framework* para metadatos en la www desarrollado por el W3C.

Capítulo 2. Propuesta de solución

3. SPARQL *Protocol and RDF Query Language* (SPARQL): lenguaje para la consulta de datos RDF, que permite hacer búsquedas sobre los recursos de la Web Semántica utilizando distintas fuentes de datos, cubren tres objetivos:

- Extraer información en forma de URIs y literales.
- Extraer sub-estructuras RDF.
- Construir nuevas estructuras RDF partiendo de resultados de consultas.

La Web de datos (*web of Linked data*) pretende facilitar el acceso, publicación y reutilización de los datos a través de la web, así como el desarrollo de aplicaciones en torno a dichos datos. No sólo se centra en la publicación y reutilización de los datos, sino también en la generación de aplicaciones que hagan uso de los mismos. Este punto es importante ya que diferencia la web de datos de la actual web, puesto que no se basa en documento sino en datos que se pueden utilizar de diferentes formas en diferentes aplicaciones (DELGADO, 2010).

El supuesto básico detrás de *Linked data* es que el valor y la utilidad de los datos aumenta a medida que se interrelaciona con otros datos.

Se puede acceder a la Web de los datos mediante navegadores *Linked data*, al igual que la web tradicional de los documentos se accede utilizando navegadores HTML.

Sin embargo, en lugar de seguir los vínculos entre las páginas HTML, los navegadores *Linked data* permiten a los usuarios navegar entre las diferentes fuentes de datos siguiendo los enlaces RDF.

Ya hay una gran cantidad de datos estructurados accesibles en la web a través de las API, como el de eBay, Amazon, Yahoo y Google Base API. En comparación con estas API, *Linked data* tiene la ventaja de proporcionar un mecanismo de acceso único y estandarizado en lugar de confiar en las interfaces y formatos de resultados. Esto permite que las fuentes de datos puedan ser:

- Más fácilmente rastreadas por los motores de búsqueda.
- Accesibles mediante navegadores de datos genéricos.
- Que permita vínculos entre los datos de diferentes fuentes de datos.

Capítulo 2. Propuesta de solución

Actualmente en la Web existe un proyecto desarrollado por el grupo de la W3C, encargado de divulgar y explicar la Web Semántica, lleva por nombre Linking Open Data y su objetivo principal es ampliar la web con una base de datos común mediante la publicación en la web de bases de datos en RDF y mediante el establecimiento de enlaces RDF entre datos de diferentes fuentes. En octubre de 2007 se contabilizaron conjuntos de datos de más de dos millones de tripletas, relacionados entre sí por medio de más de dos millones de enlaces RDF.

A continuación se muestran unas imágenes donde se exponen los conjuntos de datos que están publicados en formato *Linked data* y que están vinculados con otros conjuntos de datos en la web, además de la evolución que ha alcanzado la web a medida que las aplicaciones y bases de datos implementaron los principios propuestos por Berners-Lee, a estas imágenes se les llama comúnmente en la Web de datos la nube de *Linked data* (*Linked data cloud*).

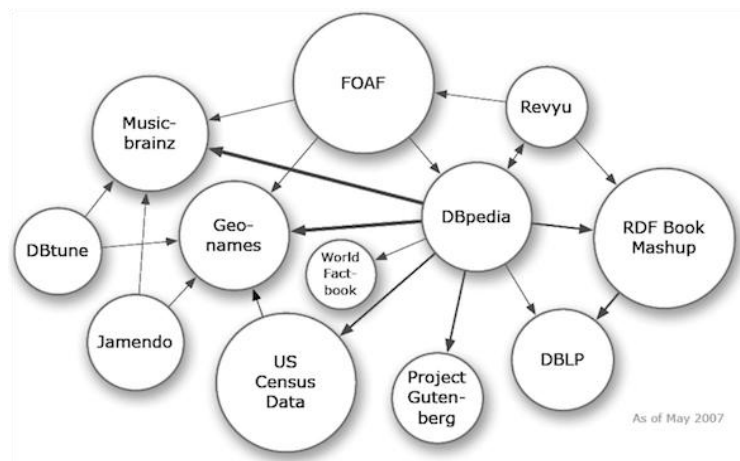


Figura 3. Nube de datos (Linking Open Data) de Mayo 2007.

En estas imágenes las flechas significan que entre estos conjuntos de datos hay al menos 50 vínculos o consultas realizadas. Entiéndase por vínculo las tripletas RDF donde el sujeto y el objeto identificados con URIs se encuentran en diferentes conjuntos de datos. El tamaño de los círculos indica el número de tripletas que hay en los conjuntos de datos.

Capítulo 2. Propuesta de solución

formato *Linked data* se utiliza un fichero de configuración de la extensión para convertir los datos contenidos en las bases de datos.

Para poder crear este fichero primero hay que realizar consultas a las bases de datos para poder recoger la información que se quiere publicar y convertirlas en tripletas RDF. Por ejemplo en la primera columna debe aparecer el identificador que se utilizará para crear las URIs de las instancias, las columnas de las tablas de la base de datos se pueden renombrar para reutilizar propiedades de vocabularios existentes como FOAF.

LH RDF: es un plugin que crea automáticamente un SIOC³¹ compatible con RDF de las publicaciones, categorías, etiquetas y autores de Wordpress.

Viene con la negociación de contenido integrado y se incorpora un atributo de enlace en la sección <head> para la detección automática.

La publicación de un RSS en RDF XML con URI dereferenceable, adecuadamente expone todos los objetos públicos de Wordpress y hacerlo en el formato SIOC lo hace de una manera estandarizada.

Es el primero de una serie de plugins para permitir ser un verdadero motor de blog semántico.

Agrega un nuevo tipo de hilo de sindicación al que se puede suscribir, por ejemplo <http://example.com/feed/lhrdf/> o <http://example.com/?feed=lhrdf> que contienen datos semánticos en formato RDF. Esto permite la producción de metadatos SIOC de su weblog (SHAW, 2012a).

LH Relationships: es un plugin que permite la creación de artículos a relaciones entre artículos. Fue construido como parte del núcleo del proyecto de LH RDF. Es el segundo de una serie de plugins que permite que Wordpress sea un motor de blog de semántico. Una vez activado creará una tabla que habilitará las relaciones entre artículos, donde se guardan las entre los artículos, un formulario en la página de edición de los mismo para entrar los datos y una API para acceder a los datos en los temas (SHAW, 2012b).

Las extensiones encontradas daban solución parcial al problema. No pudo hacerse uso de las mismas pues necesitaban conexión directa a Internet para acceder a servicios publicados que se encargan de

³¹ Comunidades en línea conectadas semánticamente/ *Semantic-Interlinked Online Communities*.

Capítulo 2. Propuesta de solución

realizar la conversión de los datos a datos semánticos, optando por desarrollar una plantilla para la exposición de los contenidos noticiosos.

Para darle solución al problema científico planteado en la investigación se incluyó una funcionalidad al CMS Wordpress desarrollada sobre una plantilla (template).

El CMS Wordpress propone una estructura propia para los sistemas web desarrollados sobre él. Trabaja con los archivos independientes, de manera que cada archivo pueda modificarse y adaptarse a las necesidades del administrador del sistema. La configuración típica de un sitio en Wordpress contiene los archivos de configuración del sistema de publicación, además de carpetas con funciones que luego serán usadas por el CMS agrupadas en archivos con extensión PHP. Estas carpetas principales son: *admin*, *includes* y *content*.

La carpeta *content* está compuesta a su vez por dos carpetas más: extensiones (*plugins*) y temas (*themes*) que precisamente será la carpeta donde se guardará la plantilla (template) sobre la que se desarrolló la funcionalidad por ser la carpeta definida por Wordpress para guardar las plantillas que conformarán el tema a utilizar en la aplicación.

En las cabeceras de los contenidos de tipo noticia se incluirá automáticamente un enlace para ser interpretado por los agentes inteligentes a incluirse próximamente en la red social; entre los elementos del enlace se encuentra el identificador de la noticia, el cual será utilizado para obtener los metadatos asociados a la misma, como son: título, autor, fecha, copyright, hora, cuerpo de la noticia, los cuales serán presentados en formato NewsML cumpliendo con el estándar XML. El contenido obtenido será luego manejado en el núcleo de la red social y puesto a disposición de otras aplicaciones para su posterior utilización.

Con esta funcionalidad se asegura el cumplimiento de los cuatro principios enunciados por Berners-Lee para caracterizar la información en *Linked data*; pudiéndose constatar en los ejemplos que se ilustran a continuación:

1. Primer principio plantea: utilizar URIs para identificar los recursos, este se pone de manifiesto cuando las descripciones de los recursos se exponen a través de la definición de URIs únicas para describir cada recurso (noticias, artículos, etc.).

```
<?php
$base = get_bloginfo('home');
$for = '/newsml/?id=';
$identificador = get_the_id();
?>
<link rel="meta" type="application/rdf+xml" title="NewsML" href="
  <?php echo $base.$for.$identificador;?>" />
```

Figura 7. Creación de la URI.

Funciones utilizadas:

- `get_bloginfo('home')`: función que permite obtener el camino hasta la URL base del periódico digital.
 - `$for = '/newsml/?id='`: ruta donde será invocada la plantilla que expondrá el contenido de la noticia en formato NewsML.
 - `$identificador = get_the_id()`: función que permite obtener el identificador de la noticia el cual será utilizado para extraer los metadatos asociados a las mismas.
2. El segundo principio plantea: aprovechar el HTTP de la URI para que la gente pueda localizar y consultar los recursos. Las descripciones de los recursos se proveen a través del protocolo HTTP para que sean consultadas por terceras aplicaciones, tomando como ejemplo el núcleo de la red social que a través de este protocolo realizará la recuperación de estos datos.
 3. Proporcionar información útil acerca del recurso cuando se consulte la URI mediante los estándares. La información de los recursos se proporcionan basándose en la ontología propuesta NewsML que realiza la descripción de los recursos usando como serialización el lenguaje XML.

Capítulo 2. Propuesta de solución

```
.NewsML>
  <Catalog Href="http://www.iptc.org/std/catalog/catalog.IptcMasterCatalog.xml"/
  <NewsEnvelope>
    <DateAndTime><?php echo mysql2date('Ymd', get_lastpostmodified('GMT'), fal
    <NewsService FormalName="Thumbnail" link-url="http://localhost/wordpress">
      http://localhost/wordpress/wp-content/themes/twentyeleven/screenshot.p
    </NewsService>
  </NewsEnvelope>
<?php foreach ($posts as $post) { ?>
  <?php setup_postdata($post); ?>
  <NewsItem LinkType="normal">
    <Identification>
      <NewsIdentifier>
        <ProviderId><?php the_author(); ?></ProviderId>
      </NewsIdentifier>
    </Identification>
  <NewsManagement>
</NewsManagement>
</Identification>
  </NewsIdentifier>
    <ProviderId><?php the_author(); ?></ProviderId>
```

Figura 8. Generación de código NewsML.

4. Incluir enlaces a otras URIs, de forma que se potencie el descubrimiento de información en la Web. En este caso el descubrimiento de información útil se potenciará al acceder a las URIs de los autores y a ciertas partes del cuerpo de la noticia, que referenciarán información útil y relacionada con el tema.

Capítulo 3. Validación de la propuesta

CAPÍTULO 3: VALIDACIÓN DE LA PROPUESTA

3.1. INTRODUCCIÓN

En este capítulo se realizará la validación de la propuesta realizada, para ello se tomarán como referencia los resultados obtenidos por algunas aplicaciones que usan este mecanismo. Estas experiencias ayudarán a demostrar si el mecanismo es factible y si se han alcanzado resultados concretos a nivel mundial. Este método no asegura que las instituciones y aplicaciones que utilizaron este mecanismo hayan logrado el éxito en el mundo digital con su utilización, pero si muestra que lograron exitosamente aplicar el mismo y que les proporcionó varias ventajas.

3.2. VALIDACIÓN DE LA PROPUESTA

Como se ha corroborado la finalidad de *Linked data* es enlazar datos desde diversas fuentes con el fin de poder generar un conocimiento más amplio con respecto a un tema, además, de que permite que los seres humanos sean capaces de compartir datos estructurados fácilmente, de manera similar a la forma como se comparten actualmente documentos en la web. En la actualidad se han hecho numerosos esfuerzos para investigar y desarrollar aplicaciones que se aprovechen de esta gran nube de datos vinculados existentes en la web. Estos esfuerzos se pueden clasificar en tres categorías:

- Navegadores de datos vinculados.
- Motores de búsqueda de datos vinculados.
- Aplicaciones con datos vinculados.

La validación de la propuesta se enmarcará en las aplicaciones con datos vinculados.

Los principios y prácticas de *Linked data* han sido adoptados por un número creciente de proveedores de datos, de información, proporcionando un resultado invaluable como la creación de un espacio global de datos en la web.

3.2.1. PRIMER CASO: BBC

Como primer caso de éxito se encuentra el portal del grupo de comunicación británica BBC, compañía de transmisión de noticias que utiliza datos vinculados internamente como una integración tecnológica de datos. La BBC utiliza tecnologías semánticas basadas en los principios de *Linked data* para algunos de los servicios que ofertan como BBC Music (para la música) y BBC Programmes (para programas).

Capítulo 3. Validación de la propuesta

Estos servicios conforman su información a partir de conjuntos de datos externos que se encuentran publicados en la web. Esta información procede de diferentes bases de datos como MusicBrainz (base de datos de música) y DBpedia (abstracción semántica de Wikipedia), y es unida a los sitios de la BBC a través de vocabularios controlados que conectan contenidos del mismo tema residentes en diferentes repositorios. Esto les ha permitido aumentar el contenido publicado en sus páginas con datos adicionales provenientes de la nube de información de Linked Open Data independientemente del dominio en el que se encuentren.

Esta gran compañía también es capaz de crear contenido semántico, mediante la identificación de contenidos de interés con URIs, por ejemplo el servicio BBC Music le asigna un identificador a cualquier artista que aparezca en los Programmes de la BBC, en cualquier evento, permitiendo la reutilización de contenido, además de facilitarle a cualquier desarrollador aprovechar estos contenidos para construir sus propias aplicaciones, puesto que sus páginas están disponibles en XML, RDF, entre otros estándares (GEORGI KOBILAROV, 2009).

The screenshot shows the BBC Music website interface. At the top, there is a navigation bar with the BBC logo, a search bar, and a 'Text only Help' link. Below this is the 'Music BETA' header with a 'QUICK FIND' search bar. The main content area features a large image of a man playing a harmonica, with a 'BBC 1XTRA Watch 1Xtra Live' video player overlay. Below the image is a 'BROWSE BY GENRE' section with a list of music genres. The 'Most Played Artists On The BBC' section displays a horizontal bar chart for the period '30 NOVEMBER - 6 DECEMBER 2009', listing artists like Dizzee Rascal, Johnny Cash, Mumford & Sons, Jason Derulo, Arctic Monkeys, Donae'o, and Jay. The bottom of the page includes 'Latest News Stories' and 'Music Radio Programmes' sections.

Figura 9. BBC Music (BBC Music).

Capítulo 3. Validación de la propuesta

3.2.2. SEGUNDO CASO: THE NEW YORK TIMES

El segundo caso de éxito es el del sistema de gestión de noticias The New York Times.

Por más de 150 años, el grupo del periódico The New York Times ha indexado cuidadosamente sus archivos. A lo largo de este proceso lograron desarrollar una enorme colección de titulares que se extienden desde la letra "A" hasta la "Z".

Desafortunadamente dicha lista de temas es una isla, es decir, tiene una capacidad exuberante y presenta limitaciones, por ejemplo, aún cuando son capaces de mostrar todos los artículos escritos sobre una persona en específico, digamos "Colbert, Stephen", las BD del periódico no pueden proveer más información, como por ejemplo que el señor nació el 13 de mayo de 1964, o que perdió el Grammy del 2008 al mejor álbum hablado a Al Gore, para poder hacer esto necesitarían asignar los temas titulares hacia otras BD en la web como Freebase y DBpedia.

En el año 2009 comenzaron con una asignación manual de más de 5.000 nombres de personas en Freebase y DBpedia. Y hoy en día cuentan con una página <http://data.nytimes.com> donde liberan esos 5.000 nombres como Linked Open Data. Con la liberación de estos datos, actualmente se puede visitar <http://data.nytimes.com/N66220017142656459133> y ver que el mencionado anteriormente "Colbert, Stephen" es equivalente al artículo de DBpedia http://dbpedia.org/resource/Stephen_Colbert y al de Freebase [http://rdf.freebase.com / RDF / en.stephen_colbert](http://rdf.freebase.com/en/stephen_colbert).

Lo más importante es que se puede acceder a <http://data.nytimes.com/N66220017142656459133.rdf> y obtener toda esta información en formato RDF para su posterior reutilización aunque también proveen información en HTML (EVAN SANDHAUS, 2009).

Para asegurar que la información publicada en ese sitio se utilizara de forma más amplia y lo más libre posible, los contenidos fueron publicados bajo la licencia Creative Commons Attribution 3.0³².

3.2.3. TERCER CASO: DBPEDIA MOBILE

El tercer caso es el sitio es DBpedia Mobile, un navegador *Linked data* de reconocimiento de ubicación, diseñado para ejecutarse en teléfonos móviles.

Basada en la actual posición GPS del dispositivo móvil, la aplicación proporciona un mashup de lugares cercanos a la posición del móvil, utilizando las bases de datos DBpedia para proveer las

³² Es una de varias licencias de derecho de autor, que permite la distribución de obras con derecho de autor.

Capítulo 3. Validación de la propuesta

ubicaciones exactas, Revyu (sitio web donde se pueden hacer críticas y posicionamientos a lugares, películas, personalidades, etc.) para proveer las críticas asociadas y el API de Flickr que provee las fotos de dichos lugares.

Además de acceder a los datos de la web, DBpedia Mobile también permite a los usuarios publicar su ubicación actual, fotos y las críticas en la web como *Linked data*, para que puedan ser reutilizados por otras aplicaciones y usuarios. En lugar de solo etiquetar las locaciones con coordenadas geográficas, también se interrelacionan con el contenido publicado en la DBpedia y por lo tanto contribuye a la riqueza global de la Web de Datos (TIM BERNERS-LEE).



Figura 10. DBpedia Mobile mostrando información sobre Berlín. (TIM BERNERS-LEE)

3.2.4. CUARTO CASO: ESTADÍSTICAS INE

El año pasado la CTIC puso a consideración de la población española una aplicación que, mediante el uso de *Linked data*, permite visualizar de manera sencilla datos de Estadísticas INE sobre equipamiento y uso de las TIC y el comercio electrónico en las empresas y hogares asturianos.

Lo más destacable de esta aplicación es que, gracias al uso de las tecnologías *Linked data*, permite combinar datos estadísticos almacenados en decenas de hojas Excel con diferentes estructuras y nomenclatura, una tarea que anteriormente resultaba tediosa y complicada.

Capítulo 3. Validación de la propuesta

Utilizando tecnologías semánticas como RDF o SPARQL se consiguió modelar, unificar y exponer los datos, haciendo posible realizar comparaciones de los indicadores estadísticos a lo largo del tiempo (evolutivos), comparaciones entre los datos asturianos y nacionales y, además, se deja la puerta abierta para, en el futuro, poder comparar los datos entre diferentes comunidades autónomas.

Como no podía ser de otra forma en un proyecto Open Data, los datos utilizados por la aplicación se han hecho públicos y son accesibles desde el portal de Reutilización de Información del Sector Público del Principado de Asturias para que cualquiera pueda acceder a ellos y utilizarlos en sus propios desarrollos (*Aplicación para Estadísticas INE*).

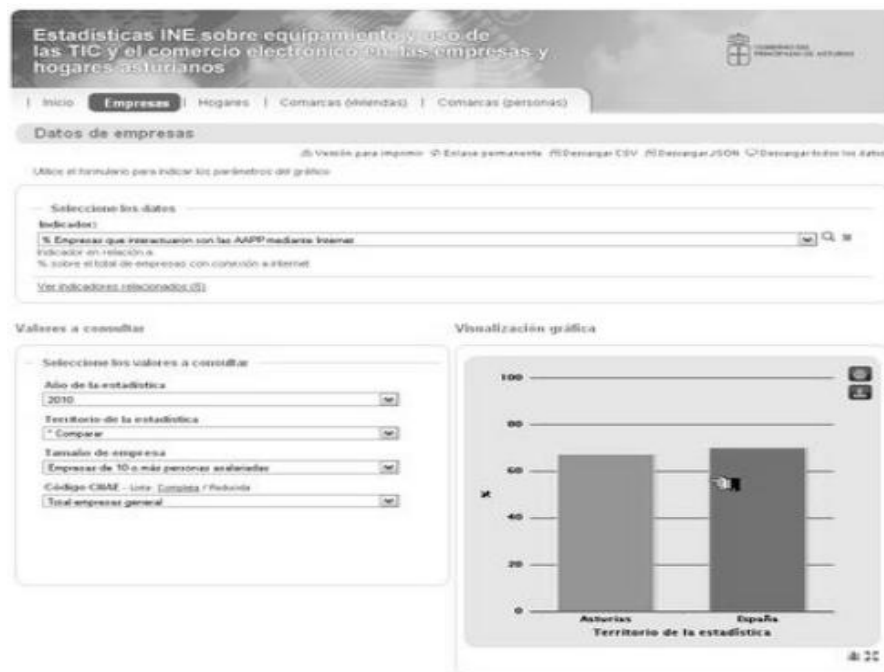


Figura 11. Aplicación para el tratamiento de datos estadísticos en España. (*Aplicación para Estadísticas INE*).

3.2.5. QUINTO CASO: GEOLINKED DATA

GeoLinked Data es una iniciativa abierta destinada al enriquecimiento de la Web de los Datos con datos geospaciales del territorio nacional español. Esta iniciativa se puso en marcha con la publicación de diversas fuentes de información geográfica, haciéndolas disponibles como bases de conocimiento RDF conforme a los principios de *Linked data*. Además, estos datos se interrelacionan con otras bases de conocimiento existentes en la iniciativa Linking Open Data. De esta manera,

Capítulo 3. Validación de la propuesta

España se suma a la iniciativa que otros países como Reino Unido y Alemania han comenzado recientemente.

Las fuentes de información geoespaciales publicadas proceden del Instituto Geográfico Nacional de España (IGN) y del Centro Nacional de Información Geográfica (CNIG). Estas fuentes se caracterizan por recoger información múltiple a diferentes escalas y por contener información de carácter multilingüe en las lenguas oficiales de España (castellano, catalán, gallego, vasco y aranés) (*GeoLinked Data*, 2010).

CONCLUSIONES

1. A través del estudio realizado en la presente investigación se demostró la necesidad de aplicar un mecanismo de integración entre el periódico digital Mella y la Red Social Universitaria.
2. El uso de tecnologías basadas en la Web Semántica permitió la estandarización de los contenidos a publicar en la red.
3. El análisis de los diferentes mecanismos de integración permitió seleccionar como el más idóneo para resolver el problema planteado *Linked data* y NewsML como estándar para la descripción de las noticias.
4. Para realizar la validación de la propuesta de solución se analizaron un conjunto de aplicaciones e instituciones que exitosamente hacen uso de este método, permitiendo corroborar la factibilidad del mismo.

RECOMENDACIONES

- La puesta en práctica del mecanismo de integración propuesto en el periódico digital universitario.
- Ampliar el campo investigativo a otros componentes que permitan realizar las mismas funcionalidades propuestas en los CMS más utilizados en la Universidad, dígase Drupal y Joomla.

Referencias Bibliográficas

REFERENCIAS BIBLIOGRÁFICAS

ACOSTA, R. M., REMEDI, R. M. *BASE DE DATOS LIBRES*. 2008. vol. 1.2,

ALVAREZ, M. A. *Una introducción meramente conceptual al potente lenguaje de script del lado del cliente*. publicado el: 01 de enero de 2001, última actualización: 01 de enero. Disponible en: <http://www.desarrolloweb.com/articulos/25.php>.

ÁLVAREZ, R. *Las primeras cosas que debes saber sobre HTML: historia, objetivos y demás conocimientos donde sentar las bases del manual*. publicado el: 24 septiembre de 2001, última actualización: 24 septiembre Disponible en: <http://desarrolloweb.com/articulos/534.php>.

AMADOR, A. P. 2008, vol. 2012, Disponible en: <http://informaticadesdeexilio.wordpress.com/2008/04/23/sistema-de-gestion-de-contenidos-wordpress/>.

Aplicación para Estadísticas INE. Disponible en: <http://datos.fundacionctic.org/asturiastic/>.

BAIL, S. *API*. publicado el: 29 enero de 2009, última actualización: 29 enero Disponible en: <http://www.cabinetng.com/media/API.pdf>.

BBC Music. Editado por: Bbc.Png.

BERNERS-LEE, T. *Linked data. Design Issues* Última actualización: 18 junio 2009. [Consultado el: 20 abril 2012 Disponible en: <http://www.w3.org/DesignIssues/LinkedData.html>].

CASTRO, L. 2012, Disponible en: <http://aprenderinternet.about.com/od/RedesSociales/g/Que-Es-Una-Red-Social.htm>.

CIBERAULA. *Una Introducción a Apache*. <http://linux.ciberaula.com>, [Consultado el: 3 de diciembre de 2011]. Disponible en: http://linux.ciberaula.com/articulo/linux_apache_intro.

CODINA, L. *Web Semántica y Sistemas de Información Documental*. 2009.

Definición de XML. 2008, Disponible en: <http://definicion.de/xml/>.

DELGADO, M. A. G. *MÉTODO PARA LA GENERACIÓN, PUBLICACIÓN Y VISUALIZACIÓN DE INFORMACIÓN EN FORMATO LINKED DATA*. Trabajo de diploma, ESCUELA SUPERIOR DE INGENIERÍA INFORMÁTICA INGENIERÍA INFORMÁTICA, 2010.

EVAN SANDHAUS, R. L. *First 5000 tags released to the linked data cloud* [Consultado el: 29 Octubre

FRESNO, M. D. *Red Social Online (una propuesta de definición)*. 28 septiembre de 2006 2010, nº Disponible en: <http://www.migueldelfresno.com/2010/09/red-social-online-una-propuesta-de.html>.

GeoLinked Data. Disponible en: <http://www.oeg-upm.net/index.php/linkeddta/73-geolinkeddta>.

GEORGI KOBILAROV, T. S., YVES RAIMOND, SILVER OLIVER, CHRIS SIZEMORE, MICHAEL SMETHURST, CHRISTIAN BIZER, ROBERT LEE. 2009, Disponible en: <http://derivadow.files.wordpress.com/2009/06/eswc2009-bbc-dbpedia-2.pdf>.

Referencias Bibliográficas

- Guía breve de CSS. 2008, Disponible en: <http://www.w3c.es/Divulgacion/GuiasBreves/HojasEstilo>.
- IBARRA, O. T. *GESTORES DE BASES DE DATOS* publicado el: 16 de febrero de 2010, última actualización: 16 de febrero. Disponible en: <http://infocomp4.blogspot.com/2010/02/gestores-de-bases-de-datos.html>.
- INTERNATIONAL DIGITAL ENTERPRISE ALLIANCE, I. I. *PRISM: Publishing Requirements for Industry Standard Metadata* publicado el: 12 marzo 2004 de 2004, última actualización: 12 marzo 2004. Disponible en: [http://www.primstandard.org/specifications/PRISM1\[1\].2.pdf](http://www.primstandard.org/specifications/PRISM1[1].2.pdf).
- IPTC, C. *News Exchange Format* [Consultado el: 05 mayo 2012 Disponible en: http://www.iptc.org/site/News_Exchange_Formats/NewsML_1/].
- LAPUENTE, M. J. L. *Hipertexto, el nuevo concepto de documento en la cultura de la imagen*. tesis doctoral, Universidad Complutense de Madrid, 2011.
- MALDONADO, D. *7 características de NetBeans 6.9.1 integrado a PHP* <http://www.elcodigok.com.ar>, [Consultado el: 3 de diciembre de 2012]. Disponible en: <http://www.elcodigok.com.ar/2010/09/7-caracteristicas-de-netbeans-6-9-1-integrado-a-php/>.
- NETBEANS. *Bienvenido a NetBeans* www.netbeans.org, [Consultado el: 3 de diciembre de 2011]. Disponible en: http://netbeans.org/index_es.html.
- PABLO CASTELLS, F. S. En *Ontologías para la web semántica*. Escuela Politécnica Superior Universidad Autónoma de Madrid. 2003.
- PHP (PHP Hypertext Pre-processor). Disponible en: http://ideocentro.com.ar/recursos.php?id=14&estudio_diseno_web_ideocentro=PHP--PHP-Hypertext-Pre-processor-.
- PRESERVATION, C. D. *Sustainability of Digital Formats Planning for Library of Congress Collections* [Consultado el: 03 mayo 2012 Disponible en: <http://www.digitalpreservation.gov/formats/fdd/fdd000014.shtml>].
- ROUSE, M. publicado el: Septiembre de 2005, última actualización: Septiembre. Disponible en: <http://searchsoa.techtarget.com/definition/XMLNews>.
- RUIZ, R. G. *TFC: XML y Web Semántica. Estudio del impacto de las aplicaciones comerciales basadas en tecnologías de Web Semántica*. Tutor: Gómez, S. A. Ingeniería Técnica Informática de Sistemas. 2008.
- SHAW, P. *LH RDF* Última actualización: 20 mayo 2012. Disponible en: <http://wordpress.org/extend/plugins/lh-rdf/>.
- SHAW, P. *LH Relationships* Última actualización: 22 abril 2012. Disponible en: <http://wordpress.org/extend/plugins/lh-relationships/>.
- TELLO, A. L. *Ontologías en la Web Semántica*. En *I Jornadas de Ingeniería Web'01*. 2001.

Referencias Bibliográficas

TIM BERNERS-LEE, C. B., TOM HEATH. *Linked data, The Story so Far*. Disponible en: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>.

TROLLTECH, J. B. *The Little Manual of API Design*. publicado el: 19 junio 2008 de 2008, última actualización: 19 junio 2008. Disponible en: <http://www.ecured.cu/index.php/API>

ZAMORA, M. 2006, Disponible en: <http://www.maestrosdelweb.com/editorial/redessociales/>

BIBLIOGRAFÍA CONSULTADA

- RSS. Disponible en: <http://www.euroresidentes.com/Diversión/Internet/rss.htm>.
- Triplify description*. 2001, Disponible en: <http://linux.softpedia.com/get/System/System-Administration/Triplify-36198.shtml>.
- ALVAREZ, M. A. *Vemos qué es RSS, rápidamente y en líneas generales. RSS es un formato de distribución de contenidos de Internet basado en XML*. [Consultado el: 13 de Mayo] Disponible en: <http://www.desarrolloweb.com/articulos/que-es-rss.html>.
- AMADOR, A. P. 2008, vol. 2012, Disponible en: <http://informaticadesdeexilio.wordpress.com/2008/04/23/sistema-de-gestion-de-contenidos-wordpress/>.
- BAIL, S. *API*. publicado el: 29 enero 2009 de 2009, última actualización: 29 enero 2009. Disponible en: <http://www.cabinetng.com/media/API.pdf>.
- BAIL, S. *API*. publicado el: 29 enero de 2009, última actualización: 29 enero Disponible en: <http://www.cabinetng.com/media/API.pdf>.
- CASTRO, L. 2012, Disponible en: <http://aprenderinternet.about.com/od/RedesSociales/g/Que-Es-Una-Red-Social.htm>.
- CODINA, L. *Web Semántica y Sistemas de Información Documental*. 2009.
- CURTO, J. publicado el: 28 noviembre de 2006, última actualización: 28 noviembre. Disponible en: <http://informationmanagement.wordpress.com/category/gestion/gestion-de-la-informacion/>.
- DELGADO, M. A. G. *MÉTODO PARA LA GENERACIÓN, PUBLICACIÓN Y VISUALIZACIÓN DE INFORMACIÓN EN FORMATO LINKED DATA*. Trabajo de diploma, ESCUELA SUPERIOR DE INGENIERÍA INFORMÁTICA INGENIERÍA INFORMÁTICA, 2010.
- DEVELOPERS, F. [Consultado el: 10 mayo 2012] Disponible en: <http://developers.facebook.com/docs/reference/api/>.
- GEORGI KOBILAROV, T. S., YVES RAIMOND, SILVER OLIVER, CHRIS SIZEMORE, MICHAEL SMETHURST, CHRISTIAN BIZER, ROBERT LEE. 2009, Disponible en: <http://derivadow.files.wordpress.com/2009/06/eswc2009-bbc-dbpedia-2.pdf>.
- IPTC, C. *News Exchange Format* [Consultado el: 05 mayo 2012] Disponible en: http://www.iptc.org/site/News_Exchange_Formats/NewsML_1/.
- LAPUENTE, M. J. L. *Hipertexto, el nuevo concepto de documento en la cultura de la imagen*. tesis doctoral, Universidad Complutense de Madrid, 2011.

- PEREZ, S. *Semantify Your Web Apps with Triplify* [Consultado el: 20 mayo 2012 Disponible en: http://www.readwriteweb.com/archives/semantify_your_web_apps_with_triplify.php.
- PRESERVATION, C. D. *Sustainability of Digital Formats Planning for Library of Congress Collections* [Consultado el: 03 mayo 2012 Disponible en: <http://www.digitalpreservation.gov/formats/fdd/fdd000014.shtml>.
- RUIZ, R. G. *TFC: XML y Web Semántica. Estudio del impacto de las aplicaciones comerciales basadas en tecnologías de Web Semántica*. Tutor: Gómez, S. A. Ingeniería Técnica Informática de Sistemas. 2008.
- TROLLTECH, J. B. *The Little Manual of API Design*. publicado el: 19 junio 2008 de 2008, última actualización: 19 junio 2008. Disponible en: <http://www.ecured.cu/index.php/API>
- VERGARA, G. 2009, Disponible en: <http://mejoratugestion.com/mejora-tu-gestion/que-es-un-sistema-de-gestion/>.
- ZAMORA, M. 2006, Disponible en: <http://www.maestrosdelweb.com/editorial/redessociales/>